Current Biology 16, 221–229, February 7, 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2005.12.034

Article

Convergent Evolution of Receptors for Protein Import into Mitochondria

Andrew J. Perry,[1] Joanne M. Hulett,[1] Vladimir A. Likić,[1] Trevor Lithgow,[1,*] and Paul R. Gooley[1,*]
[1] Department of Biochemistry and Molecular Biology
Bio21 Institute of Molecular Science and Biotechnology
University of Melbourne
Parkville, Victoria 3010
Australia

Summary

Background: Mitochondria evolved from intracellular bacterial symbionts. Establishing mitochondria as organelles required a molecular machine to import proteins across the mitochondrial outer membrane. This machinery, the TOM complex, is composed of at least seven component parts, and its creation and evolution represented a sizeable challenge. Although there is good evidence that a core TOM complex, composed of three subunits, was established in the protomitochondria, we suggest that the receptor component of the TOM complex arose later in the evolution of this machine.
Results: We have solved by nuclear magnetic resonance the structure of the presequence binding receptor from the TOM complex of the plant *Arabidopsis thaliana*. The protein fold suggests that this protein, *At*Tom20, belongs to the tetratricopeptide repeat (TPR) superfamily, but it is unusual in that it contains insertions lengthening the helices of each TPR motif. Peptide titrations map the presequence binding site to a groove of the concave surface of the receptor. In vitro functional assays and peptide titrations suggest that the plant Tom20 is functionally equivalent to fungal and animal Tom20s.
Conclusions: Comparison of the sequence and structure of Tom20 from plants and animals suggests that these two presequence binding receptors evolved from two distinct ancestral genes following the split of the animal and plant lineages. The need to bind equivalent mitochondrial targeting sequences and to make similar interactions within an equivalent protein translocation machine has driven the convergent evolution of two distinct proteins to a common structure and function.

Introduction

Mitochondria were derived some two billion years ago from endosymbiotic bacteria [1–3]. This singular event saw the gradual transfer of genes from bacteria to the host-cell nucleus, requiring that the proteins now encoded in the nucleus of the host be targeted to and assembled in the newly established organelle [2]. Several hundred nuclear-encoded proteins have now to be targeted to mitochondria. These proteins generally carry mitochondrial targeting signals that are short, basic presequences with a propensity to form a stretch of amphipathic α helix. Because targeting sequences appear to have existed as preadaptations on the bacterial proteins of the endosymbiont [4], the development of a molecular machine for protein import was the crucial hurdle to initiate mitochondrial evolution. The working model for the evolution of these molecular machines proposes that a simple channel protein in the outer membrane of the early endosymbiont existed in a host cell expressing some substrates predisposed for targeting to mitochondria [4, 5]. Comparative genomic analyses showed Tom7, Tom22, and Tom40 to be the earliest components of the TOM complex, derived before the divergence of animals and fungi from protozoans and plants [5].

Functional analysis of mitochondria from both fungi and animals show that substrate proteins are delivered to the core complex that includes Tom7, Tom22, and Tom40 by transient docking with the import receptors, Tom20 and Tom70 [6–10]. These receptor subunits work together to assist precursor protein entry into the TOM complex channel [7, 11]. Tom20 is the presequence receptor, binding directly to the mitochondrial targeting sequence in most, and likely all, precursor proteins. Tom70 assists hydrophobic regions of substrate proteins to remain in an unfolded conformation on the mitochondrial surface.

The Tom20 protein from rat has a hydrophobic surface patch that provides a binding site for the mitochondrial targeting signal in a substrate protein [12], and comparative analyses showed that Tom20 from fungi, invertebrates, and vertebrate animals have a conserved structure and likely evolved from a common ancestor [13]. Searches using refined hidden Markov models showed that no protein with an equivalent domain structure is found in plants or in the various protist species for which genome sequence is available [13]. Previous work with mitochondria isolated from plants identified a 20 kDa mitochondrial protein as a protein import receptor. Antibodies raised to the 20 kDa protein blocked protein import into plant mitochondria [14], and the 20 kDa protein copurified on two-dimensional blue native PAGE with the components of the TOM complex [15]. These critical analyses leave open two important questions: Does the 20 kDa plant protein directly bind presequences, and is there any evolutionary relationship between the plant protein and the Tom20 subunit from animals and fungi?

Evolution tends to proceed divergently, with an ancestral organism giving rise to one or more new species, and most evolutionary theory, even when viewed at the molecular level of genes and proteins, is best explained in terms of radiating evolutionary trees [16, 17]. However, there are cases in which evolution proceeds convergently, such that organisms of two distinct ancestors, each faced with an equivalent set of pressures, ultimately evolve similar structures [17]. Here, we show that the 20 kDa subunit of the plant TOM complex is

*Correspondence: t.lithgow@unimelb.edu.au (T.L.); prg@unimelb.edu.au (P.R.G.)

a presequence binding receptor and with ^{15}N HSQC titrations reveal that the plant Tom20 binds presequences in a similar manner to the yeast Tom20. The foundation for the structure of the plant presequence binding receptor is a pair of TPR segments, and the receptor domain is attached to the mitochondrial outer membrane through a flexible linker to a single transmembrane segment. All these structural features are equivalent to the structure of Tom20 in animals and fungi, but in the reverse order. Thus, the plant Tom20 must have been derived from a distinct ancestral gene. The Tom20 receptors found in animals and plants have no homologs in any of the various groups of protists, suggesting that they evolved sometime after the radiations of these eukaryote groups, and have converged to be structurally and functionally equivalent proteins.

Results

A 20 kDa Protein from *Arabidopsis thaliana* Is Functionally Equivalent to Animal and Fungal Tom20

Although a 20 kDa protein has been identified as part of the TOM complex of plant mitochondria, it had not been demonstrated to be functionally equivalent to the presequence binding Tom20 from fungi and animals. We studied, as a model for the function of the Tom20 receptor, import of proteins into mitochondria isolated from the yeast *Saccharomyces cerevisiae*. Substrate proteins, synthesized in vitro to carry ^{35}S label, were presented to the isolated mitochondria, and their import was measured over time. Preincubation of the substrate proteins with the isolated receptor domain of Tom20 from yeast (*Sc*Tom20) inhibits their ability to bind the Tom20 on the mitochondrial surface and therefore inhibits their import (Figure 1A). This is true for the matrix-targeted Su9-DHFR and the intermembrane space protein cytochrome b_2 (Cytb2): Both proteins are known to require Tom20 for import [11]. An equivalent amount of the isolated receptor domain of the 20 kDa protein from the plant *A. thaliana* (*At*Tom20; accession number: P82874) inhibits the import of these substrate proteins into yeast mitochondria. Because the model protein Su9-DHFR consists only of a short mitochondrial targeting sequence fused to a nonmitochondrial passenger domain, both *At*Tom20 and *Sc*Tom20 most likely bind the mitochondrial targeting sequence irrespective of the passenger protein.

To further test that plant and animal/fungal Tom20 receptors bind to similar regions of presequences, we incubated an ^{15}N-labeled mitochondrial targeting sequence attached to a carrier fusion protein (GB1-CoxIV) with either the plant *At*Tom20 or the fungal *Sc*Tom20 receptor to observe any chemical-shift perturbation in the GB1-CoxIV ^{15}N-HSQC spectrum (Figure 1B). When titrated with either receptor, GB1-CoxIV shows very similar patterns of amide chemical-shift perturbation, providing a simple fingerprint for comparison of binding. Several resonances of the targeting sequence broaden and disappear whereas others remain mostly unperturbed. The broadening of similar resonances strongly suggests that the equivalent part of the targeting sequence is bound by *At*Tom20 and *Sc*Tom20.

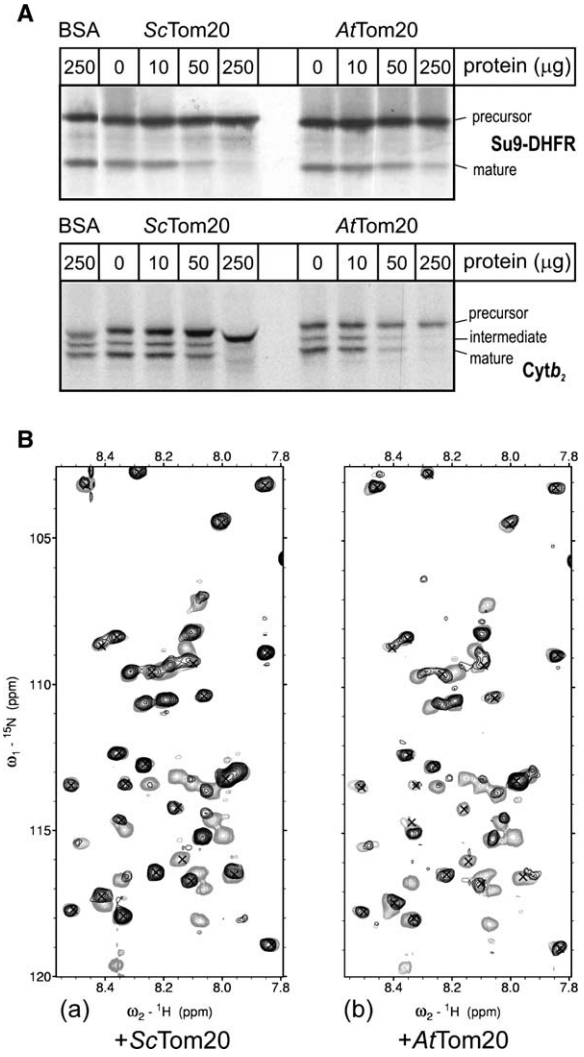

Figure 1. Both the Fungal *Sc*Tom20 and Plant *At*Tom20 Bind Mitochondrial-Targeting Sequences

(A) The cytosolic domains of *Sc*Tom20 and *At*Tom20 were purified, and the indicated amounts were added to a reaction mixture containing ^{35}S-labeled precursor protein and mitochondria isolated from yeast. Initial experiments demonstrated that these reaction conditions (7 min, 25ºC, with 25 μg protein of mitochondria) were linear with respect to time and mitochondrial amounts. Control (BSA) shows import measured in the presence of 250 μg bovine serum albumin. Import of the precursor proteins Su9-DHFR and cytochrome b_2 (Cytb2) was measured in the presence of pure recombinant *Sc*Tom20 (0, 10, 50, and 250 μg) or pure recombinant *At*Tom20 (0, 10, 50, and 250 μg). Import was terminated by the addition of the inhibitor FCCP, and the samples were subsequently treated with PMSF and SDS-PAGE loading buffer [48]. The samples were analyzed by SDS-PAGE and fluorography.

(B) ^{15}N-HSQC titrations of ^{15}N-GB1-CoxIV(1-25,C19S) with unlabeled receptor cytosolic domains. Light-gray contours are GB1-CoxIV only, whereas black contours are (a) 1:1 GB1-CoxIV:*Sc*Tom20 and (b) 1:1 GB1-CoxIV:*At*Tom20. Peaks due to the GB1 domain are marked with a cross (determined by using free GB1 samples without CoxIV targeting sequence), whereas peaks due to the CoxIV targeting sequence portion are unmarked.

Solution Structure of the Receptor Domain of a Plant Tom20

The structure of the cytosolic domain of Tom20 from rat (*Rn*Tom20) has been solved by NMR spectroscopy and

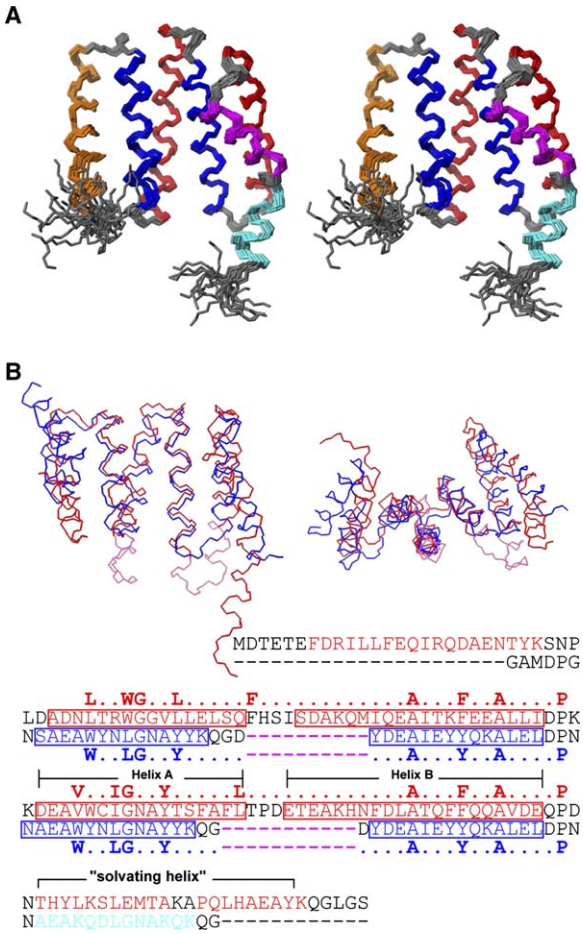

Figure 2. Structure of *At*Tom20

(A) Stereo view of 20 superimposed NMR-derived structures of *At*Tom20. Regions of helix are colored, and turn and loop regions are gray. The residue positions of the helices are α1, 7–25; α2, 31–46; α3, 50–70; α4, 75–91; α5, 95–115; α6, 120–130; and α7, 132–141. (B) Superposition of *At*Tom20 (red) with the designed TPR protein CTPR3 (blue) [19] (pdb: 1NA0) shown as a backbone trace. Alignments are based on RMS fit of equivalent Cα atoms from residues at TPR or TPR-like consensus positions. Helices in the corresponding sequence alignment are boxed, with consensus residues indicated in bold above or below their respective sequences. These residues form complimentary faces between helices. Notably, the packing of the side chains of Trp37 (of *At*Tom20) with Lys63 forms what appears to be an aromatic-cation interaction [49]; however this is only observed in the first TPR motif.

consists of five helices [12]. The first and second helices form a single tetratricopeptide repeat (TPR) fold and, together with the third helix, form the presequence binding site. Figure 2A shows that *At*Tom20 consists of seven antiparallel α helices. A conserved Pro (Pro133) in the C-terminal helical region results in two short helices punctuated by a small kink where we would otherwise expect a single long helix. N- and C-terminal residues (encompassing residues 1–6 and 142–153) are disordered and not shown in Figure 2A. Random-coil chemical shifts and low steady state $\{^1H\}$-^{15}N NOE values (data not shown) are consistent with a lack of structure for these regions.

Four of the helices (α2, α3, α4, and α5) form two TPR-like motifs. The classical TPR motif was originally defined as the 34 amino acid residue consensus [WLF]-X(2)-[LIM][GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE] [18]; however, a more extensive superfamily can be defined by using profile HMMs, such as those produced by PFAM (http://www.sanger.ac.uk/Software/Pfam/). Structurally, the TPR comprises two antiparallel α helices (A and B), arranged with interhelical angles of −149° to −162° (B and A′) and of 11° to 32° (A and A′) between the helices of succeeding TPRs [19]. The TPR-like structures of *At*Tom20 are atypical only in that they are considerably longer (43 and 44 residues, respectively). However, these structural elements of *At*Tom20 still retain the interhelical angles of a classical TPR protein. Aligning the structure of *At*Tom20 with typical TPR proteins shows that the additional residues within the TPRs are in both cases extensions of the helices rather than loop insertions (Figure 2B). The extended TPR helices in *At*Tom20 would provide similar binding surfaces, for ligands and for partner proteins, as those found in proteins composed of classical TPR helices.

Comparison of the Cytosolic Domain of the Plant Tom20s to Other Clans of TPR Proteins

Structural alignments of *At*Tom20 to representative TPR proteins allow us to determine how well residues of the consensus sequence of the TPR motif are conserved within the two TPR motifs of the plant Tom20s (Figure 2B, Figure S2 in the Supplemental Data available online). Comparison of these sequences to the classical TPR motif shows conservation of the character and spacing of the residues except for the large insertion between the helices of the TPR motif. These conserved residues play similar structural roles to those of the classical TPR motif [19]: The size and shape of hydrophobic residues separated by i + 3 or i + 4 of each helix form complimentary faces to pack and form hydrophobic core(s) within and between the TPR motifs. The first residue of the motif and the proline add stability to the turn between consecutive TPR motifs. In addition, hydrophobic residues (Phe47 and Leu91) in the large insertions are structurally conserved within the plant Tom20s, and they stabilize the packing between the N-terminal end of helix α1 and the C-terminal end of helix α2 of the first TPR motif and similarly for helix α3 and helix α4 of the first and second TPR motifs. As an aid to understanding the structural relationship of the plant Tom20 TPR and the classical TPR, a general consensus for the plant Tom20 TPR motif is proposed as: [LV]-X(2)-[WL][G]-X(2)-[LY]-X(5)-[FL]-X(13)-[A]-X(3)-[LF]-X(2)-[A]-X(4)-[P]. On the basis of this analysis, we suggest that the plant Tom20 TPR represents a new clan of the TPR superfamily.

Comparison of the Receptor Domain of Plant and Animal Tom20s

A comparison of the three-dimensional structures of the cytosolic domains of the *Rn*Tom20 and *At*Tom20 shows some similarities between the two domains (Figures 3A and 3B). A DALI search [20] using *At*Tom20 showed greater similarity to other TPR proteins than to *Rn*Tom20. Z scores greater than 9 were obtained for four TPR proteins, as well as Sec17 and one 14-3-3 protein, both of which share some structural features with the TPRs

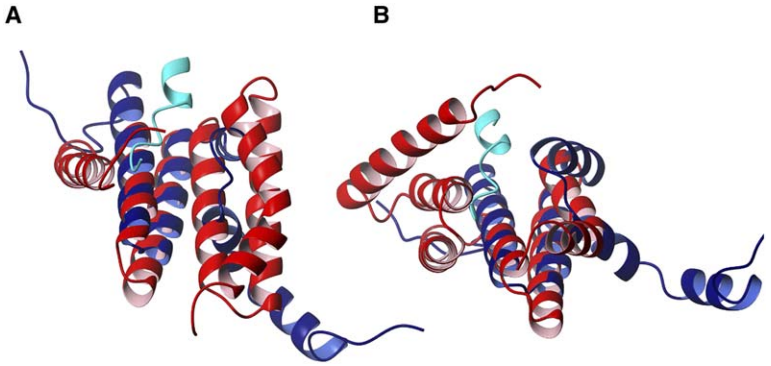

Figure 3. Structural Comparison of Plant and Animal Tom20 Proteins

Superposition of the rat *Rn*Tom20 (blue) complexed with prepeptide (cyan) (pdb:1om2, chain A) to the Cα positions of the (A) first and (B) second TPR-like repeat in *At*Tom20 (red), represented as ribbons.

[21]. *Rn*Tom20 showed lower structural similarity, with a Z score of 4.3. Thus, although both *At*Tom20 and *Rn*Tom20 are built on an equivalent structural framework, there is no evidence that they were derived from a common ancestral import receptor; instead, they are likely to be the result of the adaptation of a new function by different ancestral TPR proteins.

The Presequence Binding Site of Plant Tom20s

To identify the presequence binding site of the plant Tom20s, we titrated ^{15}N- or ^{13}C-labeled *At*Tom20 with the presequence peptides of $F_1\beta$(1–14) and GB1-CoxIV (Figure 4A). Although the magnitude of the observed chemical-shift perturbations were small, both presequence peptides affected resonances from similar residues. Mapping the most significant chemical-shift perturbations onto the structure of *At*Tom20 (Figure 4B) shows that these residues are confined to the groove of the concave surface of the receptor with the majority of residues located on the central α helices 2 and 4. Significant perturbations are also observed on α helices 1, 3, and 7. These helices are on the outer edges (helix α1 and helix α7) or the convex side (helix α3) of the domain (Figure 3). For these helices, however, residues pointing toward the concave groove are mostly perturbed. Perturbations are also observed for resonances from residues of the novel insertions of the plant Tom20 TPRs, which correspond to the turns between α helices 2 and 3 or α helices 4 and 5. Helix α5 is remote from the groove, and, consequently, no residues are perturbed. Helix α6 is on the outer edge of the groove, and it also shows no significant perturbations.

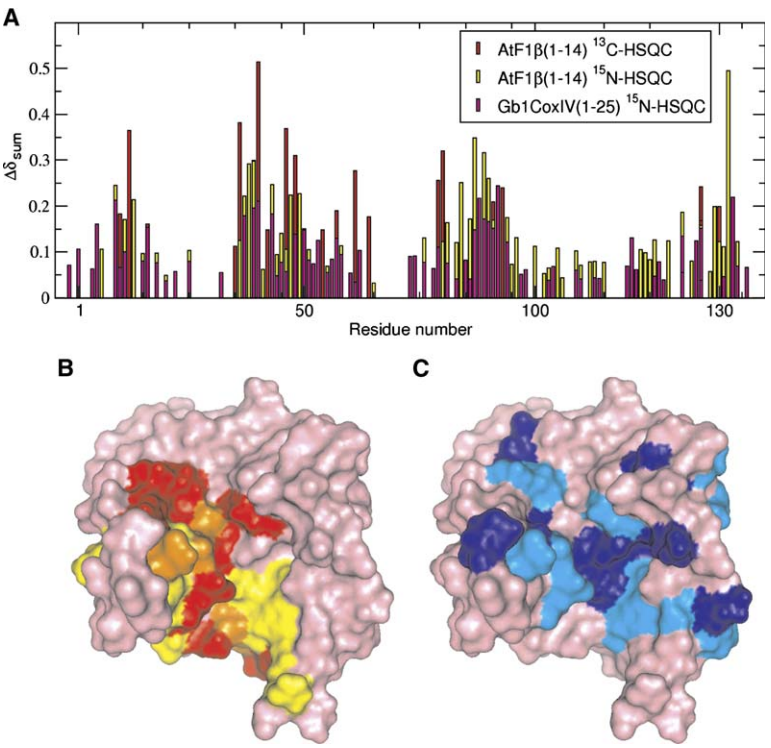

Figure 4. Presequences Interact with a Conserved Surface Groove on the *At*Tom20 Receptor

(A) A graph of chemical-shift perturbations observed upon titration of (^{13}C,^{15}N)-*At*Tom20 with the presequence peptide $F_1\beta$(1–14) (1:7, receptor:peptide) or Gb1-CoxIV (1:2, receptor:peptide). For ^{13}C/^{1}H shift changes, $\Delta\delta_{sum} = [(3 \times \Delta\delta^{H})^2 + (\Delta\delta^{C})^2]^{1/2}$, whereas for ^{15}N/^{1}H shift changes $\Delta\delta_{sum} = [(5 \times \Delta\delta^{H})^2 + (\Delta\delta^{N})^2]^{1/2}$.

(B) A molecular surface highlighting residues that display the greatest amide and aliphatic chemical-shift perturbations. Residues that show ^{13}C/^{1}H chemical-shift perturbation (^{13}C-HSQC) upon $F_1\beta$(1–14) titration are shown in red ($\Delta\delta_{sum}$ > 0.15, residues R9, I10, L12, I16, R36, G39, V40, L42, E43, Q46, H48, I50, K54, I57, I61, F64, W79, C80, L91, P93, H136, and Y140), and residues that show ^{15}N/^{1}H chemical-shift perturbation (^{15}N-HSQC) are colored yellow ($\Delta\delta_{sum}$ > 0.15, residues R9, L11, F13, W37, G38, G39, E43, F47, S49, I81, A84, T86, S87, A89, F90, L91, T92, D94, A132, H136, A139, and Q142). Regions where both amide and aliphatic groups were perturbed are colored orange. The residues that show the greatest chemical-shift perturbation trace a path through the groove on the concave surface of the receptor, indicating the likely binding site of the presequence peptide.

(C) A molecular surface displaying the conserved residues based on an alignment of 19 plant Tom20 protein sequences. The normalized Rate4Site relative evolutionary rates [22, 23] are mapped to the surface of *At*Tom20. Blue indicated values < −1.0 standard deviations from the mean rate, and cyan for values < −0.5 standard deviations from the mean, where negative rates are slower than the mean evolutionary rate. Molecular graphics were generated with PyMOL [50].

To independently identify the likely ligand binding site, we mapped residues that are most highly conserved on the surface of the plant Tom20, using comparative analysis of the Tom20 from various plant species. Evolutionary rates at each site in a multiple sequence alignment of putative plant Tom20s were estimated with Rate4Site [22, 23]. As expected, the TPR consensus and other buried hydrophobic structural residues are highly conserved (Figure 4C). The most rapidly evolving positions occur near the end of helices, in turn regions, and in disordered regions at the N termini and the C-terminal linker region. Almost all highly conserved residues on the surface of the protein face out to the concave side. Particularly notable is a conserved patch of hydrophilic residues (Asn83, Thr86, Ser87) on helix α4 and a hydrophobic (Leu42) from helix α2, which all lie at the base of the groove on the concave side of the domain (Figure 4B). Leu42, Thr86, and Ser87 are among the residues showing significant chemical-shift perturbations in the presequence-peptide-titration experiments (Figure 4A). Notably, Asn83 is equivalent to position 6 in the classical TPR motif, whereas Thr86 and Leu42 are equivalent to position 9. Residues in these two positions in several other TPR-containing proteins, including O-linked GlcNAc transferase, Tom70, and Hop, are implicated in substrate interactions [24, 25].

The Structural Elements of Plant Tom20s Are in Reverse Order Compared to Animal Tom20s

To begin to address the evolutionary relation of the Tom20 from plants and from animals, we looked to identify related proteins in genomes of a range of organisms. Clear orthologs were discovered in partially complete genome sequences from the mosses *Physcomitrella patens* and *Tortula ruralis* and the green algae *Chlamydomonas reinhardii* (Figure S1). The presence of Tom20 in *C. reinhardii* shows that the plant Tom20 had evolved at a stage before the divergence of green alga and higher plants. Notably, the plant Tom20 TPR consensus sequence is highly conserved in *C. reinhardii*, including the length of the insertions within the TPR motifs, and therefore it is likely that the structure of the algal protein is similar to that of higher plants. An antiserum raised to *At*Tom20 specifically recognizes a protein of 20 kDa in mitochondria purified from pea plants and *A. thaliana* ([26]; James Whelan, personal communication) and *C. reinhardii* (Claire Remacle, personal communication).

The plant Tom20s (Figure S1) show a series of conserved features (Figure 5A): an N-terminal disordered region with conserved acidic residues, the TPR-based receptor domain, a region rich in basic residues that is predicted to be disordered, and a C-terminal transmembrane segment. *At*Tom20 is C-terminally membrane anchored, in contrast to the N-terminally anchored Tom20 receptors found in animals (Figure 5B). Comparison of the plant and animal Tom20 transmembrane domains and the proximal cytosolic regions suggests striking structural similarities, but features within the sequences occur in reverse order. First, there are conserved glycine and aromatic residues within the region predicted to be in the bilayer of the outer mitochondrial membrane (Figure 5C). Secondly, an aspartate residue is found at the cytosolic membrane interface. Thirdly, this aspartate is followed by a region of 13 residues in which charged residues (five basic, one acidic) predominate.

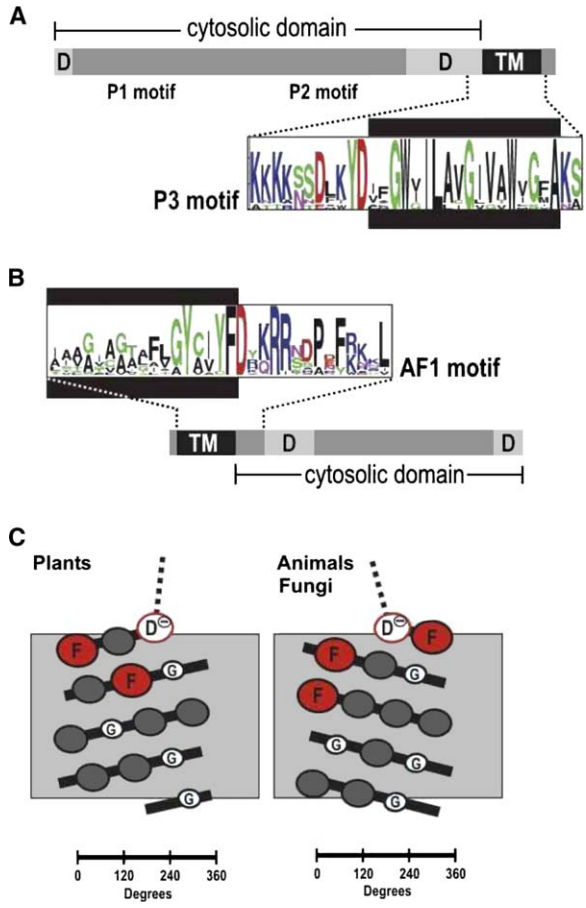

Figure 5. Conserved-Domain Features in Tom20 from Plants

(A) Motif analysis of multiple sequences of Tom20 from plants reveals a short N-terminal region and a larger internal region predicted to be highly disordered ("D"), two adjacent regions of high sequence conservation described by motifs P1 and P2 (see Figure S2), and a transmembrane segment containing several conserved residues. A Logo plot [51] of this P3 motif is shown, with the tallest representations in the motif being amino acid residues absolutely conserved across species. The black box represents the part of the motif predicted to be within the plane of the outer membrane.

(B) An equivalent representation of the Tom20 proteins from animals and fungi [13] showing the disordered regions and the residues conserved in the transmembrane AF1 motif.

(C) Helical net projection, representing in two dimensions how the transmembrane section of the AF1 (animals and fungi) and P3 (plants) motifs would be displayed at the protein-lipid interface in the mitochondrial outer membrane. Conserved residues are designated as follows: red "F" denotes large aromatic residues (Phe, Trp, and Tyr); light-blue "G" denotes small residues (Gly and Ala); and gray denotes other less-well-conserved hydrophobic residues. The dashed line represents each motif's charged region, which sits at the cytosolic surface of the outer membrane.

Hidden Markov models generated from the sequence motifs found in plant Tom20 sequences incorporate specific features in their TPR motifs and should be able to determine other ancestrally related proteins, but find no further sequences in the current UniProt database.

Discussion

The transmembrane domain and its flanking regions of the plant and animal Tom20s only show clear sequence and structural similarities when viewed in reverse, and genetic mechanisms underlying protein evolution, such as duplication, cyclic permutations, and limited insertions and deletions, could not easily result in the sequence reversal that we observe in these two proteins. We therefore hypothesize that two distinct TPR protein ancestors existed prior to the split of the lineage giving rise to plants and protozoans from that giving rise to animals and fungi. These distinct ancestral TPR proteins independently gave rise to the Tom20 in plants and the Tom20 in animals and fungi.

A current model for the evolution of the TOM complex posits that a simple, core TOM complex consisting of the channel subunit Tom40, together with Tom7 and Tom22, was established in the protomitochondria to initiate a pathway for protein import from the cytosol [9]. Mitochondrial proteins were "preadapted" for import, with features such as basic amphipathic extensions already present in the proteins of the endosymbiont [8]. These presequences mediate interactions with acidic domains of Tom22 and with surfaces of the Tom40 channel [27]. Our findings that the plant and animal lineages independently evolved presequence binding receptors to enhance protein import suggest that Tom20 was a later add-on to the original protein translocase. The same might be true for Tom70, the other receptor subunit of the TOM complex, because no Tom70 ortholog has been found in plants [28]. That the original TOM complex was "receptorless" is a reasonable proposition, given that yeast mutants lacking both Tom20 and Tom70 remain viable, but only if levels of the core subunits like Tom22 are maintained [29, 30].

Complete genome sequences are available covering the broad diversity of major eukaryotic groups, including alveolates (e.g., *Plasmodium* and ciliates), trypanosomatids (e.g., *Trypanosoma*, *Leishmania*), chromists (e.g., *Thalassiosira*), and red algae (e.g., *Cyanidioschyzon*), but the hidden Markov models used here and previously [13] did not reveal proteins from these organisms with the characteristics of either the Tom20 found in animals and fungi or the Tom20 found in plants. In addition, we recently developed a hidden Markov model that describes the Tom70 import receptor (N.C. Chan, V.A.L., and T.L., unpublished data). Deliberate attempts with these hidden Markov models to screen the sequence data from the diverse groups of eukaryotes did not identify proteins related to either Tom20 or Tom70. Although too little sequence data exists to be certain, it seems that these groups have either no receptors or have independently evolved receptors for protein import into mitochondria. A compelling rationale also exists for studying the protein import receptors in anaerobes such as *Trichomonas vaginalis* and *Giardia intestinalis* with mitochondria-like organelles (hydrogenosomes and mitosomes) [31, 32]; the characteristics of their protein import machinery promise to shed new light on the evolution of mitochondria.

Conclusions

The Tom20 in animals is anchored to the outer membrane via its N terminus, whereas the plant Tom20 is C-terminally anchored. A simple model of the transmembrane helix shows that the placement of the conserved glycine and hydrophobic residues is similar if one sequence is viewed in reverse compared to the other, and we hypothesize that this presents a similar three-dimensional surface for docking to other components of the TOM complex. Only after the endosymbiont had established, and the split of eukaryotic lineage had occurred, were each of these two ancestral TPR proteins independently recruited for the role of presequence binding receptor of the evolving TOM complex. Since that time, the need for interactions made within an equivalent protein translocation machine, the core TOM complex, but in distinct life-forms, has driven the convergent evolution of two distinct proteins to a common function.

Experimental Procedures

Protein Import Assay into Mitochondria

Whole mitochondria were isolated from *S. cerevisiae* strain W303 and purified on Nycodenz gradients as described previously [33]. Two mitochondrial precursor proteins were used for import into mitochondria: Su9-DHFR (matrix-targeting signal from subunit 9 of the mitochondrial F_1F_0-ATPase fused to dihydrofolate reductase) and cytochrome b_2 (Cytb2). Up to 250 μg of pure recombinant *At*Tom20 or *Sc*Tom20 or bovine serum albumin was combined with 4 μl of rabbit reticulocyte lysate containing in vitro translated, ^{35}S-methionine-labeled mitochondrial precursor protein and import buffer (0.6 M sorbitol, 50 mM HEPES [pH 7.4], 2 mM potassium phosphate, 25 mM KCl, 10 mM $MgCl_2$, 0.5 mM EDTA, 1 mM dithiothreitol). After 4 min at 25ºC, 25 μg of mitochondria was added to each tube (total volume 100 μL) and incubated for a further 7 min to permit the import of mitochondrial precursor protein. Import was stopped by addition of 50 μM fluorocyanide *m*-chlorophenylhydrazone (FCCP). Mitochondria and proteins associated with the mitochondrial surface were collected by centrifugation and subjected to SDS-PAGE and autoradiography.

Binding of ^{15}N-Labeled Presequence Peptides by NMR Spectroscopy

The coding sequence for cytochrome oxidase subunit IV, corresponding to amino acid residues 1–25, was PCR amplified from yeast genomic DNA, introducing a Factor Xa protease cleavage site N-terminal to the CoxIV sequence and a Cys19Ser mutation within it. The insert was ligated between the BamHI and EcoRI sites in the pGEV2 [34] vector to produce pGEV2-CoxIV. ^{15}N-GB1-CoxIV was expressed with the pGEV2-CoxIV vector in *E. coli* by using shaker-flask labeling in ^{15}N-ammonium chloride minimal medium [35] and purified by IgG-Sepharose affinity chromatography. Aliquots of concentrated, unlabeled *Sc*Tom20 or *At*Tom20 receptor (in 20 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA, pH 7.4) were added to 350 μl of 0.1 mM ^{15}N-GB1-CoxIV in the same buffer, such that the molar ratio of ^{15}N-GB1-CoxIV:receptor was 1:1 and the total volume change was less than 10%. ^{15}N-HSQC spectra were acquired at 25ºC before and after addition of the receptor. An ^{15}N-HSQC of ^{15}N-GB1, with the CoxIV peptide region removed by thrombin cleavage and repurification on IgG sepharose, was used to identify resonances arising from the GB1 domain.

Solution-Structure Determination by NMR Spectroscopy

Details of the overexpression of GST-*At*Tom20-3-His_6 (encompassing residues 1–145 of *At*Tom20-3; accession number: P82874), labeling with ^{15}N and ^{13}C, and purification are described elsewhere [36]. Samples of *At*Tom20-His_6 protein were further purified by anion exchange on a MonoQ column (Pharmacia) and concentrated into the final NMR buffer (0.5 mM and 1.0 mM *At*Tom20, in 90% H_2O/10% ^2H_2O or ^2H_2O, 20 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA, 1 mM TCEP, pH 7.4). NMR spectra were recorded at 25ºC on a Varian 600 INOVA equipped with a ^{1}H, ^{15}N, ^{13}C single z axis gradient probe. Spectral assignment of the backbone and side-chain resonances of

*At*Tom20 was achieved with standard triple-resonance experiments and is described elsewhere [36]. Distance constraints were derived from 3D-^{15}N-edited-NOESY-HSQC, 3D-^{13}C-edited-HSQC-NOESY ($^{1}H_2O$), 3D-^{13}C-edited-HSQC-NOESY ($^{2}H_2O$), and 3D-aromatic-^{13}C-edited-NOESY-HSQC ($^{2}H_2O$, carrier at 125 ppm), all with 100 ms mixing times. Spectra were processed with NMRPipe [37] and analyzed with SPARKY (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California). Stereospecific assignments of valine and leucine methyl groups were obtained by using a ^{1}H,^{13}C-HSQC spectrum recorded on ^{13}C-labeled *At*Tom20 prepared from media containing 10% $^{13}C_6$-glucose and 90% of $^{12}C_6$-glucose [38]. NOE data were assigned by using the CANDID utility of CYANA 1.0.7 [39, 40] with multiple rounds of automatic and manual assignment. A final set of 1627 NOE constraints was supplemented with 129 $^{3}J_{(HNH\alpha)}$ couplings (± 2 Hz error), 232 backbone ϕ/ψ dihedral constraints determined with TALOS [41] and 68 hydrogen-bond constraints derived from $^{1}H/^{2}H$ exchange experiments (consistent with α-helical structure from initial calculations without hydrogen-bond constraints included) (Table 1). The final ensemble of 20 structures was analyzed by PROCHECK-NMR [42], 98.9% of the residues are in allowed regions of the Ramachandran plot, and 90.2% are in the most favored regions. The molecular structure figures were generated with MOLMOL [43]. The best 20 representative structures of *At*Tom20 have been deposited at the Protein Data Bank: RCSB ID code rcsb033118 and PDB ID code 1ZU2.

Chemical-Shift Mapping of Presequence Peptide Binding to ^{15}N- or ^{13}C-Labeled AtTom20

A sample containing 0.1 mM ^{15}N-*At*Tom20 (20 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA, 1 mM TCEP, pH 7.4) was titrated with a concentrated solution of unlabeled GB1-CoxIV or the synthetic peptide corresponding to the *A. thaliana* $F_1\beta$ ATPase presequence (residues 1–14, MASRRVLSSLLRSS) (Auspep, Melbourne, Australia). ^{15}N-HSQC spectra were acquired at intermediate points up to 1:1 final protein:prepeptide molar ratio for GB1-CoxIV and 1:10 for $F_1\beta$ (1–14). An equivalent experiment acquiring ^{13}C-HSQC spectra was recorded with 0.5 mM {^{13}C,^{15}N}-*At*Tom20 titrated with the $F_1\beta$ peptide up to a final protein:prepeptide molar ratio of 1:7.

DALI Analysis

The following pdb codes, Z scores, and function of the five TPR-containing proteins with Z scores greater than 9 were found in a DALI structure similarity search [20]: 1FCH-A (Z score 12, pex5), 1A17 (12, pp5), 1IYG (11.6, unknown function), 1HXI-A (10, pex5), and 1QQE (9.7, sec17). The 14-3-3 protein 1QJA scored 9.7. *Rn*Tom20 (1OM2) scored 4.3 and was ranked 113.

Motif Elucidation, Hidden Markov Models, and BLAST Analysis

An initial set of Tom20 sequences consisted of members collected by BLAST analysis from higher plant species (*Solanum tuberosum*, *Lycopersicon esculentum*, *Gossypium hirsutum*, *Glycine max*, *Zea mays*, *Medicago trunculata*, *Latuca sativa*, *Prunus persica*, *Capsicum annum*, *Populus tremula*, *Triticum aestivum*, and four distinct sequences from *A. thaliana*). Three programs were used for the detection of motifs inherent in plant Tom20 sequences: MEME [44], ITERALIGN [45], and PROBE [46]. Although these programs employ different algorithms for motif detection, the predicted motifs overlapped extensively with minor variations in the exact beginning and end of each motif. Given a motif predicted by all three programs, the longest stretch of sequence common to all three predictions was deemed to represent a genuine motif (the only exceptions to this rule were insertions of up to two residues, because ITERALIGN and PROBE showed a tendency to break such regions into two adjacent motifs). Three major motifs were found with this procedure, are henceforth denoted P1, P2, and P3, and occurred in the order P1-P2-P3. The motifs P1 and P2 were detected in all fifteen of the initial sequences, whereas the C-terminal motif P3 was found in 11 out of 15 sequences. The four sequences lacking motif P3 represent partial sequences, unique entries in the data set where the 3′ end of the open reading frame was absent. The motif P3 encompasses the C-terminal transmembrane region and is shown schematically in Figure 5.

With the available sequences, hidden Markov models were built corresponding to motifs P1, P2, and P3. These HMMs were combined into a library and used to search the UniProt database (UniProt Release 4.4 dated March 29, 2005, based on Swiss-Prot Release 46.4 and TrEMBL release 29.4). E value cutoffs ranging from 0.001 to 0.01 were used to extract hits to each motif. For each E value, cutoff the final list of Tom20 candidates was compiled by finding the intersection of hits to individual motifs and was examined manually. The building of HMM models and UniProt database searches were performed with the package HMMER, version 2.3.2 [47].

Each orthologous sequence was used in further rounds of BLAST to confirm similarity and seek additional Tom20 candidates. At least three isoforms of Tom20 are found in *A. thaliana* and map to chromosomes 1, 3, and 5 [15]. A fourth open reading frame (accession number P82872) is probably not expressed (Jim Whelan, personal communication). The isoform used throughout this study (accession number P82874) corresponds to the product of the gene on chromosome 3. With each of these higher-plant sequences as probes, novel sequences were identified through BLAST analysis on the EST and HGS data sets at NCBI and on the moss genome data at http://www.cosmoss.org/bm/BLAST.

The various hits, including the moss sequences, were used to probe the *C. reinhardtii* version 2.0 data set at http://genome.jgi-psf.org/chlre2/chlre2.home.html. In addition sequence data from the diatom *Phaeodactylum tricornutum* were accessed through the Diatom EST database (http://avesthagen.sznbowler.com), and sequences from *Leishmania major*, *Trypanosoma bruceii*, *Theileria parva*, and *Toxoplasma gondii* were accessed through the Sanger Centre (http://www.sanger.ac.uk/Projects/Protozoa).

Multiple sequence alignments were constructed with ClustalW by using version 1.81 with default parameters (http://www.ebi.ac.uk/clustalw). The DAS server (www.sbc.su.se/~miklos/DAS/) was used to predict transmembrane segments.

Table 1. Structural Statistics for 20 Final NMR Structures

Upper-Distance Constraints	
All	1627
Intraresidue	45
Sequential	526
Medium range (< i, i + 5)	586
Long range (> i, i + 5)	470
Hydrogen bond[a]	68
Dihedral Angle Constraints	
ϕ[b]	116
Ψ[b]	116
Total number of $^{3}J_{(HNHA)}$ coupling constraints (±2 Hz error)	129
Structural Statistics	
Maximum upper-distance violation (Å)	0.30
Sum of distance violations (Å)	7.2 ± 0.23
Maximum angle violation (°)	4.53
Average CYANA target function (Å^2)	1.19 ± 0.17
Ramachandran Statistics	
Most favored regions (%)[c]	91.8
Additionally allowed regions (%)[c]	6.0
Generously allowed regions (%)[c]	1.5
Disallowed regions (%)[c]	0.7
RMSDs to Mean Structure (Ordered Residues)[d]	
Backbone atoms (Cα, C′, N)	0.44 (± 0.07)
All heavy (nonhydrogen) atoms	0.82 (± 0.06)

[a] Hydrogen-bond constraints were derived from $^{1}H/^{2}H$ exchange experiments and consistent with α-helical structure, implemented as two upper (H^N-O 2.5 Å; N-O 3.6 Å) and two lower (H^N-O 1.5 Å; N-O 2.4 Å) distance constraints for each hydrogen bond.
[b] Backbone ϕ/ψ dihedral constraints determined with TALOS [41].
[c] Determined with PROCHECK-NMR [42].
[d] Residues 7 to 141 are considered ordered, on the basis of steady state {^{1}H}-^{15}N NOE values (data not shown) and difference from random-coil chemical shifts. Values are for the ensemble of 20 best structures (selected by target function) from a total of 100 calculated.

Antiserum Recognizing AtTom20

Details of the overexpression of GST-*At*Tom20-3-His_6 (encompassing residues 1–145 of *At*Tom20-3; accession number P82874) is described elsewhere [36]. An equivalent version of the recombinant protein was produced without the GST segment, purified, and used to immunize rabbits. The antiserum was tested at 1:10,000 dilution in immunoblots and recognizes a single protein of approximately 20 kDa after SDS-PAGE of mitochondria purified from *A. thaliana*. The serum also immunoprecipitates a single protein of approximately 20 kDa from the leaves of ^{35}S-labeled plants.

Supplemental Data

Supplemental Data include figures and are available with this article online at: http://www.current-biology.com/cgi/content/full/16/3/221/DC1/.

Acknowledgments

The *At*Tom20 cDNA clone was a gift from James Whelan, Department of Biochemistry, University of Western Australia. We thank Claire Remacle, Cameron Jennings, and Marilyn Anderson for help in assessing the antiserum raised to *At*Tom20, and we thank Ross Waller and Pavel Dolezal for critical discussions. A.J.P. was supported by an Australian Postgraduate Award; J.M.H. was supported by a Dora Lush Postgraduate Research Scholarship from the National Health and Medical Research Council (ID310656). This work was supported by a grant from the Australian Research Council (T.L. and P.R.G.).

Received: November 10, 2005
Revised: December 12, 2005
Accepted: December 13, 2005
Published: February 6, 2006

References

1. Gray, M.W., Burger, G., and Lang, B.F. (1999). Mitochondrial evolution. Science *283*, 1476–1481.
2. Kurland, C.G., and Andersson, S.G. (2000). Origin and evolution of the mitochondrial proteome. Microbiol. Mol. Biol. Rev. *64*, 786–820.
3. Hedges, S.B., Blair, J.E., Venturi, M.L., and Shoe, J.L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol. Biol. *4*, 2.
4. Lucattini, R., Likić, V.A., and Lithgow, T. (2004). Bacterial proteins predisposed for targeting to mitochondria. Mol. Biol. Evol. *21*, 652–658.
5. Maćašev, D., Whelan, J., Newbigin, E., Silva-Filho, M.C., Mulhern, T.D., and Lithgow, T. (2004). Tom22′, an 8-kDa trans-site receptor in plants and protozoans, is a conserved feature of the TOM complex that appeared early in the evolution of eukaryotes. Mol. Biol. Evol. *21*, 1557–1564.
6. Neupert, W. (1997). Protein import into mitochondria. Annu. Rev. Biochem. *66*, 863–917.
7. Endo, T., and Kohda, D. (2002). Functions of outer membrane receptors in mitochondrial protein import. Biochim. Biophys. Acta *1592*, 3–14.
8. Wiedemann, N., Frazier, A.E., and Pfanner, N. (2004). The protein import machinery of mitochondria. J. Biol. Chem. *279*, 14473–14476.
9. Söllner, T., Griffiths, T.G., Pfaller, R., Pfanner, N., and Neupert, W. (1989). MOM19, an import receptor for mitochondrial precursor proteins. Cell *59*, 1061–1070.
10. Ramage, L., Junne, T., Hahne, K., Lithgow, T., and Schatz, G. (1993). Functional cooperation of mitochondrial protein import receptors in yeast. EMBO J. *12*, 4115–4123.
11. Lithgow, T., Glick, B.S., and Schatz, G. (1995). The protein import receptor of mitochondria. Trends Biochem. Sci. *20*, 98–101.
12. Abe, Y., Shodai, T., Muto, T., Mihara, K., Torii, H., Nishikawa, S.-I., Endo, T., and Kohda, D. (2000). Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. Cell *100*, 551–560.
13. Likić, V.A., Perry, A., Hulett, J., Traven, A., Waller, R., Keeling, P.J., Curran, S.P., Koehler, C.M., Gooley, P.R., and Lithgow, T. (2005). Patterns that define the four domains conserved in known and novel isoforms of the protein import receptor Tom20. J. Mol. Biol. *347*, 81–93.
14. Heins, L., and Schmitz, U.K. (1996). A receptor for protein import into potato mitochondria. Plant J. *9*, 829–839.
15. Werhahn, W., Niemeyer, A., Jänsch, L., Kruft, V., Schmitz, U.K., and Braun, H.-P. (2001). Purification and characterization of the preprotein translocase of the outer mitochondrial membrane from Arabidopsis. Identification of multiple forms of TOM20. Plant Physiol. *125*, 943–954.
16. Darwin, C. (1859). On the Origin of the Species by Natural Selection (London: John Murray).
17. Doolittle, R.F. (1994). Convergent evolution: The need to be explicit. Trends Biochem. Sci. *19*, 15–18.
18. Sikorski, R.S., Boguski, M.S., Goebl, M., and Hieter, P.A. (1990). A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. Cell *60*, 307–317.
19. Main, E.R.G., Xiong, Y., Cocco, M.J., D'Andrea, L., and Regan, L. (2003). Design of stable alpha-helical arrays from an idealized TPR motif. Structure *11*, 497–508.
20. Holm, L., and Sander, C. (1996). Mapping the protein universe. Science *273*, 595–603.
21. Das, K., Cohen, P.T.W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: Implications for TPR-mediated protein-protein interactions. EMBO J. *17*, 1192–1199.
22. Pupko, T., Bell, R.E., Mayrose, I., Glaser, F., and Ben-Tal, N. (2002). Rate4Site: An algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics *18*, S71–S77.
23. Mayrose, I., Graur, D., Ben-Tal, N., and Pupko, T. (2004). Comparison of site-specific rate-inference methods for protein sequences: Empirical Bayesian methods are superior. Mol. Biol. Evol. *21*, 1781–1791.
24. Jinek, M., Rehwindel, J., Lazarus, B.D., Izaurralde, E., Hanover, J.A., and Conti, E. (2004). The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. Nat. Struct. Mol. Biol. *11*, 1001–1007. Published online September 12, 2004. 10.1038/nsmb833.
25. Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F.U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: Critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell *101*, 199–210.
26. Taylor, N.L., Rudhe, C., Hulett, J.M., Lithgow, T., Glaser, E., Day, D.A., Millar, A.H., and Whelan, J. (2003). Environmental stresses inhibit and stimulate different protein import pathways in plant mitochondria. FEBS Lett. *547*, 125–130.
27. Truscott, K.N., Brandner, K., and Pfanner, N. (2003). Mechanisms of protein import into mitochondria. Curr. Biol. *13*, R326–R337.
28. Lister, R., Hulett, J.M., Lithgow, T., and Whelan, J. (2005). Protein import into mitochondria: Origins and functions today. Mol. Membr. Biol. *22*, 87–100.
29. Lithgow, T., Junne, T., Wachter, C., and Schatz, G. (1994a). Yeast mitochondria lacking the two import receptors Mas20p and Mas70p can efficiently and specifically import precursor proteins. J. Biol. Chem. *269*, 15325–15330.
30. Lithgow, T., Junne, T., Suda, K., Gratzer, S., and Schatz, G. (1994b). The mitochondrial outer membrane protein Mas22p is essential for protein import and viability of yeast. Proc. Natl. Acad. Sci. USA *91*, 11973–11977.
31. van der Giezen, M., and Tovar, J. (2005). Degenerate mitochondria. EMBO Rep. *6*, 525–530.
32. Martin, W. (2005). The missing link between hydrogenosomes and mitochondria. Trends Microbiol. *13*, 457–459.
33. Daum, G., Gasser, S.M., and Schatz, G. (1982). Import of proteins into mitochondria. Energy-dependent, two-step processing of the intermembrane space enzyme cytochrome b2 by isolated yeast mitochondria. J. Biol. Chem. *257*, 13075–13080.
34. Huth, J.R., Bewley, C.A., Jackson, B.M., Hinnebusch, A.G., Clore, G.M., and Gronenborn, A.M. (1997). Design of an

expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. Protein Sci. *6*, 2359–2364.
35. Marley, J., Lu, M., and Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. J. Biomol. NMR *20*, 71–75.
36. Perry, A.J., Hulett, J.M., Lithgow, T., and Gooley, P.R. (2005). (1)H, (13)C and (15)N resonance assignments of the cytosolic domain of Tom20 from *Arabidopsis thaliana*. J. Biomol. NMR *33*, 198.
37. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.
38. Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuthrich, K. (1989). Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. Biochemistry *28*, 7510–7516.
39. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. *273*, 283–298.
40. Herrmann, T., Güntert, P., and Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. *319*, 209–227.
41. Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR *13*, 289–302.
42. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR *8*, 477–486.
43. Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. J. Mol. Graph. *14*, 51–55.
44. Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology *2*, 28–36.
45. Brocchieri, L., and Karlin, S. (1998). A symmetric-iterated multiple alignment of protein sequences. J. Mol. Biol. *276*, 249–264.
46. Neuwald, A.F., Liu, J.S., Lipman, D.J., and Lawrence, C.E. (1997). Extracting protein alignment models from the sequence database. Nucleic Acids Res. *25*, 1665–1677.
47. Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755–763.
48. Gabriel, K., Egan, B., and Lithgow, T. (2003). Tom40, the import channel of the mitochondrial outer membrane, plays an active role in sorting imported proteins. EMBO J. *22*, 2380–2386.
49. Tatko, C.D., and Waters, M.L. (2003). The geometry and efficacy of cation-π interactions in a diagonal position of a designed beta-hairpin. Protein Sci. *12*, 2443–2452.
50. Delano, W.L. The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA, (http://www.pymol.org).
51. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Res. *14*, 1188–1190.

Accession Numbers

RCSB ID code rcsb033118 and PDB ID code 1ZU2 are being reported for the first time. The ID codes are for the same entry and are to the Protein Data Bank.